心理科学进展 2023, Vol. 31, No. 9, 1742–1755
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2023.01742

从“拟人归因”到“联盟建立”：人与聊天机器人关系对参与度的影响

磨 然 [1] 方建东 [1,2,3#] 常保瑞 [1,2,3]

([1]广西师范大学教育学部心理学系，桂林 541006)
([2]广西师范大学教育区块链与智能技术教育部重点实验室，桂林 541004)
([3]广西高校人文社会科学重点研究基地广西民族教育发展研究中心，桂林 541006)

摘 要 随着人工智能(Artificial Intelligence, AI)技术的迅猛发展，AI 聊天机器人可模拟人类指导以改善在线自助干预(Internet-based Self-help Interventions, ISIs)中用户的参与度及疗效。然而，学界对聊天机器人作用机制的探索尚处初期阶段。因此，为加深对这一问题的理性认识，文章基于人机关系的视角提出了适应 ISIs 情境的理论模型：聊天机器人可与用户经历拟人归因、功利性价值判断、发展依恋关系、建立数字治疗联盟(Digital Therapeutic Alliance, DTA)这 4 个阶段来逐步发展人与聊天机器人关系(Human−Chatbot Relationships, HCRs)，并通过 HCRs 提高用户参与度。未来研究可继续丰富 HCRs 的相关理论并检验其内在机制，基于 HCRs 理论来设计聊天机器人，深入考察影响 HCRs 的额外变量，统一参与度的操作定义并开发适合的参与度测量工具。

关键词 聊天机器人，参与度，人与聊天机器人关系
分类号 B849

1 引言

随着移动互联网的快速发展，基于传统心理咨询/心理治疗技术但更灵活、经济、便捷的在线自助干预(Internet-based Self-help Interventions, ISIs)也应运而生(Mrazek et al., 2019)。近年来，ISIs 的可行、有效性得到了较多证据支持(Izzaty et al., 2021; Johansson et al., 2021; Sun et al., 2021; Taylor et al., 2021; Weisel et al., 2019)，但用户的高脱落率及低参与度仍是 ISIs 当下要面临的巨大挑战(Taylor et al., 2021)。所谓参与度(Engagement)，指的是用户体验或用户参与 ISIs 内容的程度(Christensen et al., 2009; de Geest & Sabaté, 2003; Leeuwerik et al., 2019)，其与疗效息息相关(Asaeikheybari et al., 2021; Cavanagh et al., 2018; Karyotaki et al., 2017; Puls et al., 2020; Tetley et al., 2011)。尽管有人类支持可提振用户参与度(Baumeister et al., 2014; Richards & Richardsosn, 2012; Spijkerman et al., 2016)，但实际的人力投入也会限制 ISIs 的即时性、可拓展性以及成本效益。可喜的是，基于人工智能(Artificial Intelligence, AI)技术的聊天机器人(Chatbot)可嵌入 ISIs 中模拟人类指导，其相较于传统 ISIs 程序更能促进用户参与度及疗效(Perski et al., 2019; Provoost et al., 2020b; Vaidyam et al., 2019)。然而，国外的该领域研究仍处在对 ISIs 聊天机器人可行性、有效性进行检验的阶段，国内则鲜见相关研究，学界对 ISIs 聊天机器人的作用机制仍知之甚少(Bassi et al., 2022; Gabrielli et al., 2021; He et al., 2022; Skjuve et al., 2022b)。鉴于此，有必要加深对这一问题的理性认识，使聊天机器人的设计更具针对性以进一步促进 ISIs 的效果。故文章将通过人与聊天机器人关系(Human−Chatbot Relationships, HCRs)的视角来解释 ISIs 聊

收稿日期：2022-11-14
#为并列第一作者。
通信作者：常保瑞, E-mail: bdjs2004@163.com

天机器人的作用机制，深入讨论 ISIs 情境下 HCRs 的发展过程及其对用户参与度产生的影响，以期为后续研究提供参考。

2　HCRs 的发展过程及其对用户参与度的影响

从人机交互(Human−Computer Interaction, HCI)的角度来看，用户参与度低是由于交互体验缺乏所致，因此传统 ISIs 程序的静态交互不利于提振用户的参与积极性(Doherty et al., 2012)。反之，具有自然语言会话能力的聊天机器人能使用户主动参与而非被动接受信息，其形态从单一的聊天头像到虚拟化身(Avatar)，是一种更为积极活跃的社会行动者(Appel et al., 2012; Doherty et al., 2012; Go & Sundar, 2019; Ly et al., 2017)。此外，基于言语或非言语的会话行为是关系萌生的必要条件(Bickmore & Picard, 2005)，因此，聊天机器人将有希望与用户建立、发展 HCRs，并通过关系纽带来提升用户参与度。

鉴于 HCRs 的发展和人与人之间的关系(Human−Human Relationships, HHRs)发展有相似之处，一些 HHRs 理论也可用于解释 HCRs (Hendriks et al., 2020; Schuetzler et al., 2020)。例如，Skjuve 等人(2021a)为理解 HCRs 的发展过程而在广受验证的人际理论——社会渗透理论(Social Penetration Theory, SPT)的基础上建立了解释 HCRs 发展的三阶段模型：第一阶段，探索阶段，用户因对隐私或安全等因素的顾虑而对聊天机器人持谨慎态度，体现在自我暴露较少，且信息高度肤浅；第二阶段，情感阶段，用户对聊天机器人的功利性价值进行判断，并发展与聊天机器人的依恋关系，其交互频率与暴露深度也因而得到促进；第三阶段，稳定阶段，用户保持与聊天机器人的依恋关系，与聊天机器人的交互也成为日常生活的一部分，但自我暴露降低，并更倾向于分享日常事件。随后，Skjuve 等人(2022b)又进行了为期 12 周的纵向研究验证了此模型的解释力，并揭示了 HCRs 渐进式发展的特性。然而，此模型虽具新颖性，但迁移至 ISIs 情境仍存在着几个问题：一是此模型并未考虑人机交互之初的认知加工过程，而直接讨论了关系发展部分；二是此模型并未解释每个 HCRs 阶段下的心理机制；三是此模型更侧重于理解亲密关系的发展，若将该模型迁移至 ISIs 中，还需进一步考量心理咨询/心理治疗情境的特殊性，并调整 HCRs 的发展方向。因此，本文将在 Skjuve 等人(2021a)的基础上结合人机交互及心理学领域的理论，将 HCRs 模型进一步完整化、具体化，并使之适应数字心理咨询/心理治疗。

2.1　阶段 1：拟人归因

拟人化(Anthropomorphism)，即人类把似人特征(如形象、言语等)、动机、意图或情感赋予非人对象(如无生命的物体、动物)的认知过程(Epley et al., 2007)，具有启发式(Heuristics)的特点(Tversky & Kahnemann, 1974)。一般而言，当用户与聊天机器人初次接触时拟人化即被启动，用户会因其似人的外形或是会话能力而无意识地将其当作另一个人来对待(后简称“拟人归因”)，基于此，用户会更倾向于以人际交往策略与之交互(Nass et al., 1994)。随着交互频率增多，用户可能也会将更深入的特征，如动机、意图、情感等逐步归因于它们(许丽颖　等, 2017)，而拟人化的加强也将促进 HCRs 的发展(Pentina et al., 2023)。因此，拟人归因可能是发展 HCRs 的关键起点，而理解其内在机制将有助于促进 HCRs 的发展。

首先，用户会在内在动机的驱动下对聊天机器人作拟人归因。基于 Reeves 和 Nass (1996)提出的媒体等同理论(The Media Equation, TME)，“媒体无异于现实生活”、“人们会对媒体产生社会的、本能的反应，即便他们意识到这并不合乎理性，甚至认为自己不可能作如此反应”。而媒体等同理论的重要研究范式——计算机是社会行动者范式(Computers Are Social Actors, CASA)也认为用户倾向于直觉地关注计算机呈现的人类线索(Human Cues)，如其输出的文本、语音、语言风格、可以与人互动的特性等，并忽略计算机呈现的工具性线索，从而自然而然地将其视为与人一般的社会行动者，并以一定的社会规则(如偏见、礼貌、互惠等)对其产生社会化的反应(如信任、喜爱等)(Nass et al., 1994)。然而，用户为何会作无意识反应？基于 Epley 等人(2007)的三因素理论，拟人化由诱发主体知识(Elicited Agent Knowledge)、效能动机(Effectance Motivation)以及社会动机(Sociality Motivation)三个协同作用的因素构成。因此，在人机交互之初，用户更偏重于对非人对象的初级线索(形象、言语等)进行拟人的推理、归

因, 其目的在于为效能动机、社会动机的满足提供良好的认知基础。

其次, ISIs 情境的特殊性将促使用户产生积极的感知。基于社会线索减少理论(Reduced Social Cues, RSC), 在以计算机为媒介的通讯中用户接收到的社会线索(Social Cues)会因网络带宽的限制而减少, 这些有限的社会线索还会因为代偿效应而受到强化, 用户的心理状态也因此更容易被影响(Tanis & Postmes, 2003), 并体验到感知流畅性(Perceptual Fluency), 即一种主观上感受到的愉悦、放松(Labroo et al., 2008)。新近且广受检验的综合理论——超人际交流理论(Hyperpersonal Interaction, HI)也认为, 由于人机交互过程中社会线索的减少, 用户将更倾向于将聊天机器人过度理想化而忽视其技术缺陷, 并使用更多的印象管理策略来增进“好感” (Walther, 1996)。

最后, 用户的社会存在感(Social Presence)将得到增强, 并影响其参与度。所谓社会存在感, 即个体与另一个人在一起的感觉, 包括了共同存在(Copresence)、心理参与(Psychological Involvement)、行为参与(Behavioral Engagement)三个维度(Biocca et al., 2003)。由于用户受其内在动机及有限社会线索的影响, 其“社会存在感启发式” (Social Presence Heuristic)将得到良好激发(Sundar et al., 2008), 这也是 HCRs 发展之初, 聊天机器人影响用户参与度的核心要素。有研究指出, 被试的社会存在感被聊天机器人激活后, 其行为意向(Behavioral Intention)将显著增强(Mozafari et al., 2021)。而在 Blut 等人(2021)的元分析中也发现, 社会存在感中介了拟人化对行为意向的影响。基于计划行为理论(Theory of Planned Behavior, TPB), 行为意向是影响用户内在动机、预测用户实际参与行为的关键因素(Ajzen, 2012)。此外, 若被试感到与聊天机器人互动就像与真人互动一样生动, 那么信任感将更容易建立, 且参与度也会因此提高(Brendel et al., 2022; Hassanein & Head, 2005; Lee et al., 2021)。

综上所述, 良好的拟人归因(形象、言语等初级线索)是 HCRs 发展的前提, 用户会因此如对待人类一般与聊天机器人进行类社会互动(Parasocial Interaction), 并为效能、社会动机的满足提供良好的认知基础。随着 HCRs 的发展, 拟人化将逐渐以动机驱动为主, 拟人归因为辅, 目的是使得个体需求能得到更好的满足。因此, 在效能、社会动机的推动下, 用户将强化有限的社会线索, 进而使得社会存在感提升并影响其参与度, 而拟人归因也将更为深入(动机、意图、情感等高级线索), 并进一步促进 HCRs 发展。

2.2 阶段 2：功利性价值判断

功利性价值(Utilitarian Value), 即一项产品对用户功利性需要的满足程度, 如信息获取、效率提升、问题解决等(Choi & Drumwright, 2021)。在 HCRs 发展初期, 用户也倾向于对聊天机器人的功利性价值进行判断, 进而确定其自身的实际需要能否得到满足。具体而言, 聊天机器人是否能根据用户现阶段的期望而展现出其实际作用, 如准确且个性化的心理健康信息、流畅且自然的对话能力、精准且智慧的上下文理解能力、丰富且优质的技能服务等, 将影响用户对聊天机器人的接受度。因此, 探究功利性价值判断的机制, 将有助于具体化聊天机器人的功能设计并使 HCRs 得到促进。

首先, 用户倾向将聊天机器人定位为“工具”并关注其实用性。一方面, 拟人归因虽有利于营造良好的前期印象, 但由于用户的信任感尚未建立, 因此, 他们在与聊天机器人交互时也不会作较深入的自我暴露, 这也将导致此阶段的 HCRs 仍是肤浅的(Skjuve et al., 2021a)。同时, 有研究指出人类实际上对聊天机器人持有刻板印象, 认为它们尽管在“智能”上具有挑战人类的能力, 但其本质上并不具备情感能力。原因在于, 人类倾向将聊天机器人的反应识别为程序运算的结果, 而非“自发的”, 这也阻碍了人类与之作更深入的情感交互(Wirtz et al., 2018)。另一方面, 基于“使用和满足框架” (The Uses and Gratifications Framework; Rubin, 1983), 用户是积极、理性且目标导向的, 因此, 他们会试图寻找并选择那些能够满足他们特定需求(如互动、功用、娱乐、信息获取、社会地位等)的媒体产品, 这也是用户在 HCRs 发展之初更关注聊天机器人实际作用的原因。

其次, 功利性价值判断将影响用户的主观态度。基于广受验证的技术接受模型(Technology Acceptance Model, TAM), 用户对技术的功利性需求可归纳为有用性和易用性两方面, 它们将决定用户对技术功利性价值的态度, 并影响用户的行为意向(Legris et al., 2003)。例如, 在 Kamita 等

人(Kamita et al., 2019)的 ISIs 研究中, 聊天机器人在有用性、易用性上的得分均显著高于传统的 Web 程序, 而聊天机器人组被试(N = 15)也表现出更高的参与积极性, 并在行为意向上的得分显著高于 Web 程序组被试(N = 12)。与之类似, Park 和 Kim (2023)也发现, 感知有用性能正向预测被试与心理聊天机器人进行社会交互的意愿。而技术接受模型的修正——期望证实模型(Expectation-Confirmation Model, ECM)则进一步提出了用户评估功利性价值的可能机制, 即用户可能会将其使用产品前的期望与使用后的感知有用性进行对比, 以确定其期望是否得到证实, 而期望的证实与否将决定其满意度(Bhattacherjee, 2001)。例如, Dhiman 和 Jamwal(2023)基于期望证实模型来探究被试持续使用聊天机器人的原因, 研究结果显示, 被试对聊天机器人的感知有用性及使用后的期望证实对其满意度有显著影响。在 Xie 等人(2022)的研究中, 相较于技术性(Technology)、娱乐性(Hedonic)以及社交性(Social), 功利性(Utilitarian)是用户满意度的最强预测因素。

最后, 用户的后续参与行为也将受上述两种因素的影响。一是行为意向, 其作为个体采取某项行动的意愿强度, 将能直接影响用户的参与动机及实际参与行为(Ajzen, 2012)。若用户的行为意向水平足够高, 他们往往也会倾向于与聊天机器人交互并参与 ISIs 课程。反之, 若用户的参与动机不稳定, 则很难保证其后续的参与行为(Alfonsson et al., 2017)。二是用户满意度, 其作为更综合的评价指标, 除了能直接影响用户的参与动机外也能影响忠诚度, 是用户持续与聊天机器人交互的重要影响因素(Cheng & Jiang, 2020)。例如, Zhu 等人(2022)发现, ISIs 聊天机器人的功利性价值(个性化的信息呈现)能显著改善用户满意度, 而用户满意度也与其持续使用聊天机器人的意愿呈显著正相关。然而, 在 Liu 等人(2022)为期 16 周的 ISIs 研究中, 聊天机器人组被试的参与度随时间推移呈下降趋势。对此, 研究者认为聊天机器人除了存在技术缺陷之外, 其不能呈现有用且令被试满意的内容也是关键的原因。

总的来说, 此阶段是用户对聊天机器人的探索阶段, 用户可能会对聊天机器人持有刻板印象, 认为其仅是一个不具备情感能力的工具。在此基础之上, 通过使用和满足框架可推知, 聊天机器人需先呈现出其作为一个有效工具的价值, 在可用性、易用性、期望证实等方面证明其重要性, 并促进用户参与度。此时, 通过 Epley 等人(2007)的三因素理论可进一步推知, 随着用户与聊天机器人互动频率的增加, 他们对聊天机器人的认可度、熟悉度、确定感将得到提高。由于这一过程满足了用户对陌生事物理解、预测和掌控的需要(效能动机), 因此, 拟人化将得到增强并间接地提升用户好感度, HCRs 也将进一步发展。

2.3　阶段 3: 发展依恋关系

依恋(Attachment), 是指个体在婴孩时期与其他重要对象所发展出的一种长久且稳固的情感纽带(Bowlby, 1988), 这种经验会影响个体归因并在成长过程逐渐内化为独特的依恋风格(Attachment Styles), 并作为今后与朋友、亲人、恋人或个人所有物建立关系的模板(Bartholomew & Horowitz, 1991)。其中, 影响个体归因的关键在于其所体验到的安全感的水平, 个体每经历一次人际交互, 都会将过往的依恋风格代入其中, 并评估所获得的安全感, 进而作出影响关系发展的行为(接近或回避) (Adams et al., 2018)。近年来, 依恋理论也开始在传播学领域中被广泛应用, 以解释个体对非人对象(如宠物、品牌、虚拟人物等)情感依恋的发展以及维系关系的动机(Bauer & Woodward, 2007; Pedeliento et al., 2016; Wanser et al., 2019; Xie & Pentina, 2022)。随着 HCRs 进一步发展, 聊天机器人作为一种活跃的社会行动者, 也很有可能与用户发展出依恋关系。

首先, 人类的认知、情感加工机制有助于人机依恋的产生。例如, 有研究发现用户在与具有似人特征, 特别是呈现了关系线索(如幽默和同理心)的聊天机器人持续交互时, 他们对聊天机器人的认知推理可驱动他们对聊天机器人情感的感知, 反过来, 这种情感感知也会影响他们的认知判断(Lee, S. et al., 2020; Sánchez-Franco et al., 2021; Spatola & Wudarczyk, 2021)。在 Beck (1995)的认知模型中, 认知与情感也相互影响。此外, 基于心智感知理论(Mind Perception Theory), 用户实际上会综合认知、情感两个维度的感知对聊天机器人做拟人化的信息加工, 并形成感知心智的归因推理(Waytz et al., 2010), 进而影响其与聊天机器人互动的意愿(Blut et al., 2021)。而著名的双加工理论(Dual-Process Theory)也指出, 个体的认知过

程可分为理性的认知部分(如功利价值感知)和非理性的情感部分(如情绪价值感知) (Stanovich & West, 2000)。因此，认知和情感在人机交互的过程中相辅相成并共同塑造用户体验，并促成聊天机器人与用户的依恋关系(Abdulrahman & Richards, 2021; Bickmore et al., 2005; Choi & Drumwright, 2021; Pentina et al., 2023; ter Stal et al., 2020)。

其次，聊天机器人也可具备发展依恋关系的关键条件。一方面，聊天机器人除了更可靠、更容易被控制外，它们还能发挥“安全基地”和“避风港”的功能，通过满足个体的精神需求或是缓解其痛苦来使个体产生一定程度的依恋(Rabb et al., 2022)。例如，在 Zhou 等人(2020)的介绍中，聊天机器人“微软小冰”以其独特的情商系统使其能在交流、情感和社会归属感等多方面满足用户的精神需求，并与用户成功地建了依恋关系，进而促进用户长期且持续的活跃。而 Xie 和 Pentina (2022)研究结果也表明，当被试在痛苦和缺乏陪伴之际，若其感知到聊天机器人 Replika 的理解，并得到适当的情感支持、鼓励，他们就有可能对 Replika 产生依恋。另一方面，聊天机器人具有自然语言对话的天然优势。因此，它们的“表露”也能促进人机交互的频率及用户的自我暴露，并使用户体验到被接纳感、亲密感以及排遣孤独等持续的内在奖励，进而使依恋得到发展(Skjuve et al., 2021a, 2022b)。在社会渗透理论中，更多的信息传递及逐步的自我暴露是信任的表现，而信任也被视为关系发展的关键先决条件(Altman & Taylor, 1973)。例如，在 Kang 和 Gratch (2014)的研究中，被试在与深层自我表露的聊天机器人互动时，也会回应以更多的自我暴露，并对聊天机器人更信任、更亲密。随后，Lee, Y. C.等人(2020)的研究也发现，深层自我表露的聊天机器人比浅层自我表露及无自我表露的聊天机器人更能促进用户的自我暴露，同时也与部分被试成功地建立了依恋关系。

最后，依恋关系的发展将促使用户更长期地参与 ISIs。由上可知，拟人归因及功利性价值判断虽为 HCRs 发展提供了良好的认知基础，并促进了社会存在感及人机交互的频率，但它们均为短期视角。一方面，若只注重认知方面的影响，用户对聊天机器人的拟人化仅会停留在浅层的拟人归因(如外在形象、言语上的似人性)。另一方面，聊天机器人若越高效、实用，它们的定位则越会偏向于一个高效的“工具”而非一个具备感受能力且值得信赖的“伙伴”，这也将导致用户与它们的联结逐渐弱化，参与度也会因而下降。在心智感知理论中，若个体认为类人物体缺失情感感受能力，他们就会否认其“人性”，并拒绝与之平等交流(Waytz et al., 2010)。反之，寻求社会接触、社会联系、社会认可等社会、情感需求不但是个体在关系中最基本的诉求，也是促使其更长期、稳定地参与 ISIs 的重要因素(Epley et al., 2007)。因此，若要使 HCRs 得到稳定地增强，用户需更积极地参与人机交互，而聊天机器人也需要有所“作为”，为用户提供“情感价值”。如此，用户在社会动机的推动下才可能将聊天机器人进一步拟人化，把更深层的特征(如动机、意图、情感等)逐步归因于它们(许丽颖 等, 2017)，并与它们建立依恋关系来使自身的情感需求得到更持续、充分地满足(Xie & Pentina, 2022; 赵欣 等, 2012), HCRs 的发展也会因拟人化的加强而逐渐走向深入(Epley et al., 2007; Pentina et al., 2023)，并进一步提高用户的参与质量。

综上所述，拟人归因及对功利性价值判断均是影响用户参与 ISIs 的短期因素，若要使用户在更长的周期内保持活跃，情感因素将发挥更为重要的作用。因此，用户在社会动机的驱动下会将聊天机器人进一步拟人化并与之建立依恋纽带，在他们体验到更多积极情绪的同时, HCRs 也将趋向深入——从“工具”逐渐过渡至“伙伴”。若用户对聊天机器人的依恋转移至对 ISIs 任务的依恋，他们将更倾向于积极、持续地参与其中(McGonagle et al., 2021)，如此，将聊天机器人用于心理咨询/心理治疗方面的目标也开始得以体现。

2.4 阶段 4：建立数字治疗联盟

治疗联盟(Therapeutic Alliance, TA)，即咨询中来访者与咨询师之间的合作关系(朱旭, 江光荣, 2011)，是治疗效果的稳健预测因素(Flückiger et al., 2018)。有一系列研究发现，TA 不仅限于人与人之间，人类也能无意识地与虚拟程序建立联盟关系，而这种在 ISIs 情境中建立的联盟关系，被称之为数字治疗联盟(Digital Therapeutic Alliance, DTA)，也即人类与程序间的“合作”关系(Berger, 2017; D’Alfonso et al., 2020; Darcy et al., 2021; Heim et al., 2018)。有研究发现，DTA 不但与 TA 具有概念一致性(Conceptual Invariance)，而且

也能预测疗效的改善(Luo et al., 2022)。此外, TA 中的情感纽带维度与安全依恋相似且高度相关(Mallinckrodt & Jeong, 2015), 而不安全依恋, 特别是回避型依恋, 则是导致 TA 破裂的关键(McGonagle et al., 2021)。换至 ISIs 情境中, Hertlein 和 Twist (2018)也发现, 回避型用户可能会低活跃、低参与(排斥), 焦虑型用户可能会过度使用程序(依赖)。因此, 在依恋关系的基础上, 通过 DTA 的视角理解 HCRs 的进一步发展将更适应 ISIs 情境。

首先, 建立、发展 DTA 将更利于 ISIs 目标的达成。第一, 在实际的心理咨询中, 咨询师和来访者间应避免双重关系, 在 ISIs 中也不应例外。第二, 建立 DTA 的意义并不在于提升用户对聊天机器人的接受度、满意度, 而是为了更好地激励用户作出积极改变。第三, 建立 DTA 能使 ISIs 的过程更公开透明, 具体而言, 聊天机器人需与用户就如何实现某些目标以及如何落地而达成协议, 并对其使用的技术及隐私协议等信息开诚布公(Law et al., 2022)。由此可见, DTA 是一种经由深思熟虑的且目的明确的人机关系模式。鉴于 ISIs 的目标是使用户的心理问题得以改善, 因此, 此阶段应调节 HCRs 的发展方向, 使聊天机器人与用户建立 DTA、发展合作关系, 而非其他依恋关系(如朋友、伴侣)。

其次, 以前三阶段为基础, 建立 DTA 将更为容易。基于过往的 TA 定义, 无论是经典的三维结构(任务一致、目标一致、情感纽带) (Bordin, 1979), 亦或是四维结构(开放性、信任、合作关系、情感纽带) (Agnew-Davies et al., 1998), 都包含了认知、情感两大成分：认知成分, 指对治疗目标和任务的认可; 情感成分, 即积极的情感联结或个人依恋, 如相互信任、喜爱、尊重、关心、坦率等(朱旭, 江光荣, 2011)。而在 ISIs 中, 认知、情感也同样被认为是促进 DTA 的关键途径(D’Alfonso et al., 2020; Tong et al., 2022)。一方面, 用户在拟人归因后会无意识地把聊天机器人当作社会行动者, 并经由理性的认知过程(功利性价值判断)形成聊天机器人是否满足其需要的认知评估, 进而有助于“共识”的达成。另一方面, 随着 HCRs 的发展, 用户将对聊天机器人产生情感依恋, 进而影响情感纽带的建立。基于此, HCRs 的阶段性发展将为 DTA 的建立奠定良好基础。

最后, 通过促进 DTA 发展, 用户参与度也将进一步提振。在促进 DTA 方面, 当个体在需要关心、支持时与一个充满爱心的对象进行互动, 其安全感将得到加强, 并体验到被爱和被照顾的感觉。若这种影响能得到长期、稳定地重复, 个体的不安全依恋可能会逐渐转变为安全依恋(Bowlby, 1988; Mikulincer & Shaver, 2020; Nanjappa et al., 2014)。因此, 可通过增强安全感的交互来启动个体对安全依恋的感知, 以此强化情感纽带进而促进 DTA 发展。而对于具体的交互设计, 可参考咨询师与来访者发展 TA 的策略, 为聊天机器人设计专门的关系线索(磨然 等, 2023)。例如, 在传统的心理咨询中, 咨询师若善解人意、真诚一致并无条件积极地关注来访者, 咨访关系即可迅速、健康地发展(Rogers, 1957)。同样, Skjuve 等人(2021a)认为, 聊天机器人若要与用户建立、发展 DTA, 最关键的也是以接纳、理解、不评判等关系线索来给予用户情感支持。对此, Bell 等人(2019)也强调到, 理解、共情能力的缺乏将使得聊天机器人无法建立有效心理治疗所需的关系强度。总之, 聊天机器人可先模拟咨询师身份来建立专业、可靠的初步印象, 而后则需依托其良好的功能体验来得到用户的认可并提升复用率, 进一步则是在与用户交互的过程中兼顾咨询伦理规范, 并不断呈现强化情感纽带的关系线索(如友好、尊重、不判断、倾听、鼓励、真诚、共情、信任、自我表露等) (磨然 等, 2023)。如此, 则可在 HCRs 逐渐发展的过程中将依恋关系调节为更有利于 ISIs 目标的联盟合作关系。当 DTA 建立后, 用户的自我保护将减少, 自我暴露将更深入并更愿意合作, 其参与度及治疗效果也因此得到积极的改善(Heim et al., 2018; Liu et al., 2022; Provoost, 2021a)。例如, Goldberg 等人(2021)发现全自动的正念干预程序能够与被试建立 DTA, DTA 不但对参与度有显著的预测作用, 且干预第 3、4 周的 DTA 水平也预测了抑郁、焦虑的改善。而在 Rodrigues 等人(2021)的随机对照实验中, 被试同样与聊天机器人建立了 DTA, 而 DTA 也正向地预测了被试的参与度。

综上所述, HCRs 发展对 ISIs 中用户参与度或有重要影响, 若要用户具有较高的参与度, 实际上也要求了用户在更长的 ISIs 生命周期中留存且活跃, 这对 ISIs 程序的要求无疑也更高, 是该领域目前亟待解决的一个难题。基于此, 文章结合

了人机交互及心理学领域的成熟理论对 HCRs 的发展框架作出了进一步的完善，并提出一个适应 ISIs 情境的理论模型(图 1)。

3 现存问题及未来展望

3.1 HCRs 的相关理论缺乏，内在机制尚未明晰

在早期的研究中，尽管 Bickmore 和 Picard (2005)对聊天机器人所呈现的关系线索(如幽默和同理心)对 HCRs 的发展提供了重要启发，但他们并未构建理论来解释 HCRs 的变化。直至目前，学界中关于 HCRs 发展的理论仍未达成共识，研究者们对于 HCRs 如何启动、发展、加强，乃至于这种关系对人类的影响仍知之甚少，这也表明了在该领域中发展新理论的空间仍然较大(Muresan & Pohl, 2019; Skjuve et al., 2021a, 2022b)。由于 HCRs 的发展可能与 HHRs 的发展有相似之处，故解释 HHRs 发展的现有理论可作为理解 HCRs 的起点。因此，未来的研究可基于一些成熟的 HHRs 理论如：社会交换理论(Social Exchange Theories; Emerson, 1976)、人际投资模型(The Investment Model of Personal Relationships; Rusbult et al., 1994)、承诺信任理论(Commitment-Trust Theory; Morgan & Hunt, 1994)来更深入地理解 HCRs 的发展过程。此外, HCRs 不同阶段对人类心理影响的内在机制复杂，因此，人类线索所带来的影响在每个关系阶段的重要性或有所差异。例如，在 HCRs 发展的早期，聊天机器人的视觉、言语线索对用户满意度有重要影响(Kim et al., 2021)，而在 HCRs 发展的中后期，非言语线索、关系线索可能就占据了更重要的地位。因此，未来的研究可基于客观数据驱动的数字心理测量学(Digital Psychometrics; Latynov & Shepeleva, 2020)以及更长期的纵向研究对 HCRs 的不同阶段作严格地检验，进而得出行之有效的设计。

3.2 缺乏对聊天机器人线索的考量

首先，目前多数 ISIs 研究使用了高度面向任务型的聊天机器人，且对人类线索的设计欠缺考量，研究者们往往只区别聊天机器人有无对因变量的影响，而忽略了人类线索的重要性(Chong et al., 2021)，这将会导致结果的可比性较差并影响研究的可重复性。因此，未来的研究可汇报聊天机器人所选用的人类线索，并讨论线索与因变量之间的关系。其次，当前仅有少数研究探讨了人类线索与因变量的关系，但这些线索的设计仍缺乏理论支撑(Rapp et al., 2021)。尽管 HHRs 理论是理解 HCRs 发展的重要视角，但哪些促进 HHRs 发展的因素在 HCRs 中起作用且更为重要，HHRs 发展与 HCRs 发展有何异同，研究者们仍知之甚少。因此，未来的研究需在理论基础上设计人类线索，并以更严格的随机对照试验来检验不同线索的贡献。再次，在实际应用中，数字心理健康有多种场景，如智能导诊、情感陪伴、心理咨询、心理治疗等。研究者需意识到，不同场景中所需 HCRs 的程度有所不同。例如，智能导诊场景中用户生命周期很短，对效率的诉求相对更高，因此，拟人归因及功利性价值判断在 HCRs 中扮演了更为重要的角色。在未来的研究中，研究者可对当前的应用场景进行评估，并专门设计人类线索。最后，不同文化背景、性别、年龄段、人格特质、学历、收入水平、症状的用户对聊天机器人所呈

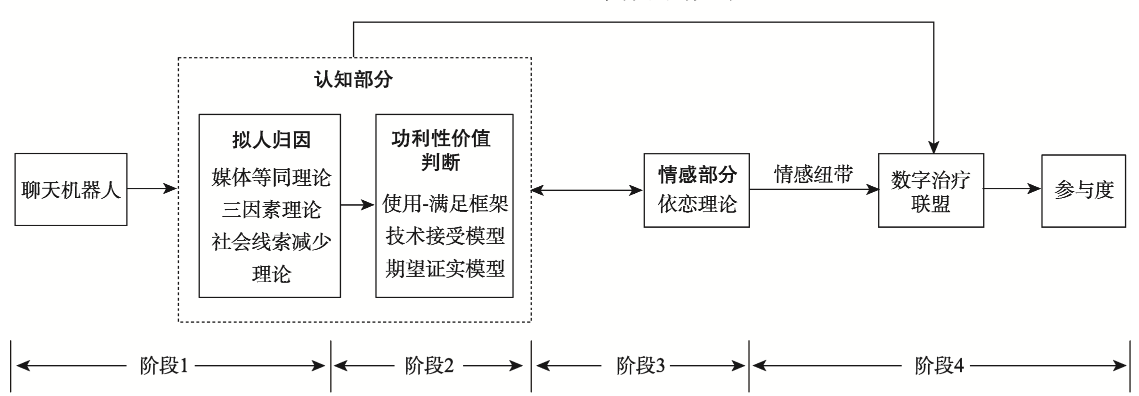


图 1 HCRs 促进参与度的理论模型(注：箭头表示可能的因果关系方向)

现的不同身份及对应的人类线索可能会产生不同的感知(Nißen et al., 2022)。因此, 未来还需根据不同场景中的用户需求及用户特征, 进一步检验能够达到更优用户体验的身份及人类线索组合。

3.3　参与度的测量与报告尚无标准

虽然, 已有一些较为初步的证据支持聊天机器人进行心理干预的可行性及有效性, 但这些研究对参与度的评估标准仍缺乏共识。首先, 研究者经常将“依从性” (Adherence)与“参与度”(Engagement)的概念混用(Beintner et al., 2019; Eysenbach et al., 2011)。然而, “依从性”实际被更多地应用于临床之中, 强调的是一种医患关系。相较于此, ISIs中的多数用户并非患者, 故文章选择“参与度”一词作为描述用户使用ISIs程度的总括术语, 这不单意味着咨访关系的平等, 且用户也需积极地参与干预。其次, 研究者对参与度的报告方法仍存在较大差异, 多数研究并未报告参与度与治疗效果的关系(Beintner et al., 2019; Vaidyam et al., 2019), 在ISIs研究中, 忽视参与度对治疗效果影响的评估, 可能会导致对干预效果的低估, 同时也影响了研究间的可比性。再次, 目前在有关ISIs的研究中, 多数研究者所使用的参与度评估指标单一且无理论支撑, 例如仅使用“完成练习的数量”这一指标, 但实际可用于评价参与度的指标十分丰富(Lederman & D’Alfonso, 2019)。因此, 未来的研究可在理论基础上统一、丰富参与度的评价指标(Beintner et al., 2019)。最后, 绝大多数ISIs研究仍过度依赖自我报告法, 但这种方法可能会导致参与度被高估(Flett et al., 2019)。因此, 在未来的研究中, 研究者可将程序后台数据、可穿戴设备采集的生物数据、人口学变量等客观数据, 与被试自我报告的主观数据进行综合分析, 以此来更全面地理解参与度这一指标。

3.4　影响HCRs的额外变量有待探究

除聊天机器人的人类线索之外, 在实际的ISIs中仍有较多因素会对研究结果造成干扰。第一, 产品性能。聊天机器人在开放域中稳定性较差, 除了时有发生的技术错误外(Jang et al., 2021), 重复且不自然的对话体验也一直被用户所诟病(Fulmer et al., 2018), 这些因素将阻碍HCRs发展。此外, 有研究发现, 相比单独的功能, ISIs项目整体发挥了更重要的作用(Berger et al., 2014), 因此究竟是聊天机器人亦或是ISIs整体带来了关键的改善仍未清晰。第二, 隐私性。在开发任何一款ISIs程序时, 大量隐私数据很可能会存在被侵犯的风险(McGreevey et al., 2020)。有研究指出, 若被试认为他们被存储在ISIs程序中的数据不够安全的话, 他们的参与度大概率会降低, 甚至会直接脱落(Proudfoot et al., 2010)。第三, 新奇效应(Novelty Effect)。用户对新鲜事物的好奇心可能会促进其短期内的积极性, 进而使其参与度被高估(Croes & Antheunis, 2020; Fryer et al., 2017; Nadarzynski et al., 2019)。第四, 恐怖谷效应。基于恐怖谷理论(Theory of The Uncanny Valley), 当类人物体逼真至一定程度, 人类对它们的喜爱亦有可能会转为厌恶(Mori et al., 2012; Song & Shin, 2022)。鉴于AI语言模型发展迅速, 聊天机器人(如ChatGPT)的仿真能力也日新月异(Aydın & Karaarslan, 2022; Elkins & Chun, 2020), 因此未来的HCRs研究也应注意此效应的影响。第五, 人类线索间的交互作用。例如, 聊天机器人使用不同的身份(机器人身份或是人类身份)将会影响用户的期望, 并因此影响其他人类线索(如外形、自然语言对话能力)的效应, 这也提示了人类线索之间可能存在交互作用, 不同线索组合对结果的影响可能也不同(Go & Sundar, 2019)。综上所述, 未来的研究需客观评估聊天机器人的作用, 并注意控制额外变量的影响, 以提升研究结果的可靠性。

参考文献

磨然, 方作之, 方建东. (2023). 如何建立聊天机器人与用户间的数字治疗联盟：关系线索的作用. *心理科学进展*, *31*(4), 669−683.

许丽颖, 喻丰, 邬家骅, 韩婷婷, 赵靓. (2017). 拟人化：从“它”到“他”. *心理科学进展*, *25*(11), 1942−1954.

赵欣, 周密, 于玲玲, 刘倩. (2012). 基于情感依恋视角的虚拟社区持续使用模型构建———超越认知判断范式. *工程管理科技前沿*, *31*(5), 14−20.

朱旭, 江光荣. (2011). 工作同盟的概念. *中国临床心理学杂志*, *19*(2), 275−280.

Abdulrahman, A., & Richards, D. (2021). In search of embodied conversational and explainable agents for health behaviour change and adherence. *Multimodal Technologies and Interaction*, *5*(9), 56. https://doi.org/10.3390/mti5090056

Adams, G. C., Wrath, A. J., & Meng, X. (2018). The relationship between adult attachment and mental health care utilization: A systematic review. *Canadian Journal of Psychiatry*, *63*(10), 651−660.

Agnew-Davies, R., Stiles, W. B., Hardy, G. E., Barkham, M., & Shapiro, D. A. (1998). Alliance structure assessed by the Agnew Relationship Measure (ARM). *British Journal of Clinical Psychology*, *37*(2), 155−172.

Ajzen, I. (2012). Martin Fishbein's legacy: The reasoned action approach. *The Annals of the American Academy of Political and Social Science*, *640*(1), 11−27.

Alfonsson, S., Johansson, K., Uddling, J., & Hursti, T. (2017). Differences in motivation and adherence to a prescribed assignment after face-to-face and online psychoeducation: An experimental study. *BMC Psychology*, *5*(1), 1−13.

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. New York, NY: Holt, Rinehart, & Winston.

Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, *2012*, 324694. https://doi.org/10.1155/2012/324694

Asaeikheybari, G., Hooper, M. W., & Huang, M. C. (2021). A context-adaptive smoking cessation system using videos. *Smart Health*, *19*, e100148. https://doi.org/10.1016/j.smhl.2020.100148

Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT generated literature review : Digital twin in healthcare. In Ö. Aydın (Ed.), *Emerging computer technologies 2* (pp. 22−31). İzmir Akademi Dernegi.

Bassi, G., Giuliano, C., Perinelli, A., Forti, S., Gabrielli, S., & Salcuni, S. (2022). A virtual coach (motibot) for supporting healthy coping strategies among adults with diabetes: Proof-of-concept study. *JMIR Human Factors*, *9*(1), e32211. https://doi.org/10.2196/32211

Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology*, *61*(2), 226−224.

Bauer, A., & Woodward, L. (2007). People and their pets: A relational perspective on interpersonal complementarity and attachment in companion animal owners. *Society & Animals*, *15*(2), 169−189.

Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2014). The impact of guidance on Internet-based mental health interventions: A systematic review. *Internet Interventions*, *1*(4), 205−215.

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York, NY: Guilford press.

Beintner, I., Vollert, B., Zarski, A. C., Bolinski, F., Musiat, P., Görlich, D., … Jacobi, C. (2019). Adherence reporting in randomized controlled trials examining manualized multisession online interventions: Systematic review of practices and proposal for reporting standards. *Journal of Medical Internet Research*, *21*(8), 1−10.

Bell, S., Wood, C., & Sarkar, A. (2019). Perceptions of chatbots in therapy. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1−6). New York, NY: Association for Computing Machinery.

Berger, T. (2017). The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. *Psychotherapy Research*, *27*(5), 511−524.

Berger, T., Boettcher, J., & Caspar, F. (2014). Internet-based guided self-help for several anxiety disorders: A randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy*, *51*(2), 207−219.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, *25*(3), 351−370.

Bickmore, T. W., Gruber, A., & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, *59*(1), 21−30.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *Association for Computing Machinery Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293−327.

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & Virtual Environments*, *12*(5), 456−480.

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632−658.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, *16*(3), 252−260.

Bowlby, J. (1988). *A secure base: Clinical applications of attachment theory*. London, UK: Routledge.

Brendel, A. B., Greve, M., Riquel, J., & Science, D. (2022, May). *" Is It COVID or a cold ?" An investigation of the role of social presence, trust, and persuasiveness for users' intention to comply with COVID-19 chatbots*. Paper presented at European Conference on Information Systems 2022, Timisoara, Romania.

Cavanagh, K., Churchard, A., O'Hanlon, P., Mundy, T., Votolato, P., Jones, F., … Strauss, C. (2018). A randomised controlled trial of a brief online mindfulness-based intervention in a non-clinical population: Replication and extension. *Mindfulness*, *9*(4), 1191−1205.

Cheng, Y., & Jiang, H. (2020). How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use. *Journal of Broadcasting and Electronic Media*, *64*(4), 592−614.

Choi, T. R., & Drumwright, M. E. (2021). "OK, Google, why do I use you?" Motivations, post-consumption evaluations, and perceptions of voice AI assistants. *Telematics and Informatics*, *62*, 101628. https://doi.org/10.1016/j.tele.2021.101628

Chong, T., Yu, T., Keeling, D. I., & de Ruyter, K. (2021). AI-chatbots on the services frontline addressing the challenges and opportunities of agency. *Journal of Retailing and Consumer Services*, *63*, 102735. https://doi.org/10.1016/j.jretconser.2021.102735

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression. *Journal of Medical Internet Research*, *11*(2), 1−16.

Croes, E. A. J., & Antheunis, M. L. (2020). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, *38*(1), 279−300.

D'Alfonso, S., Lederman, R., Bucci, S., & Berry, K. (2020). The digital therapeutic alliance and human-computer interaction. *JMIR Mental Health*, *7*(12), e21895. https://doi.org/10.2196/21895

Darcy, A., Daniels, J., Salinger, D., Wicks, P., Robinson, A., & Darcy, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. *JMIR Formative Research*, *5*(5), e27868. https://doi.org/10.2196/27868

de Geest, S., & Sabaté, E. (2003). Adherence to long-term therapies: Evidence for action. *European Journal of Cardiovascular Nursing*, *2*(4), 323−323.

Dhiman, N., & Jamwal, M. (2023). Tourists' post-adoption continuance intentions of chatbots: Integrating task−technology fit model and expectation−confirmation theory. *Foresight*, *25*(2), 209−224. https://doi.org/10.1108/FS-10-2021-0207

Doherty, G., Coyle, D., & Sharry, J. (2012). Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. *In Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems* (pp. 1421−1430). New York, NY: Association for Computing Machinery.

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a writer's turing test? *Journal of Cultural Analytics*, *5*(2), e17212. https://doi.org/10.22148/001c.17212

Emerson, R. M. (1976). Social exchange theory. *Annual Review of Sociology*, *2*(1), 335−362.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864−886.

Eysenbach, G., Stoner, S., Drozd, F., Blankers, M., Crutzen, R., Tait, R., … Bewick, B. (2011). ConSORT-eHealth: Improving and standardizing evaluation reports of web-based and mobile health interventions. *Journal of Medical Internet Research*, *13*(4), e126. https://doi.org/10.2196/jmir.1923

Flett, J. A. M., Fletcher, B. D., Riordan, B. C., Patterson, T., Hayne, H., & Conner, T. S. (2019). The peril of self-reported adherence in digital interventions: A brief example. *Internet Interventions*, *18*, e100267. https://doi.org/10.1016/j.invent.2019.100267

Flückiger, C., Del, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, *55*(4), 316−340.

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, *75*, 461−468.

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, *5*(4), e64. https://doi.org/10.2196/mental.9782

Gabrielli, S., Rizzi, S., Bassi, G., Carbone, S., Maimone, R., Marchesoni, M., & Forti, S. (2021). Engagement and effectiveness of a healthy-coping intervention via chatbot for university students during the COVID-19 pandemic: Mixed methods proof-of-concept study. *JMIR mHealth and uHealth*, *9*(5), 1−14.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304−316.

Goldberg, S. B., Baldwin, S. A., Riordan, K. M., Torous, J., Dahl, C. J., Davidson, R. J., & Hirshberg, M. J. (2021). Alliance with an unguided smartphone app: Validation of the digital working alliance inventory. *Assessment*, *29*(6), 1331−1345.

Hassanein, K., & Head, M. (2005). The impact of infusing social presence in the web interface: An investigation across product types. *International Journal of Electronic Commerce*, *10*(2), 31−55.

He, L., Basar, E., Wiers, R. W., Antheunis, M. L., & Krahmer, E. (2022). Can chatbots help to motivate smoking cessation? A study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. *BMC Public Health*, *22*(1), 726.

Heim, E., Rötger, A., Lorenz, N., & Maercker, A. (2018).

Working alliance with an avatar: How far can we go with internet interventions? *Internet Interventions*, *11*, 41−46.

Hendriks, F., Ou, C. X. J., Amiri, A. K., & Bockting, S. (2020). The power of computer-mediated communication theories in explaining the effect of chatbot introduction on user experience. *In Proceedings of the 53 Hawaii International Conference on System Sciences* (pp. 271−278). Honolulu: The Hawaii International Conference on System Sciences 2020.

Hertlein, K. M., & Twist, M. L. C. (2018). Attachment to technology: The missing link. *Journal of Couple & Relationship Therapy*, *17*(1), 2−6.

Izzaty, R. E., Astuti, B., & Cholimah, N. (2021). Evaluation of an online cognitive behavioural therapy, mindfulness meditation, and yoga (CBT-MY) intervention for posttraumatic stress disorder: A single arm clinical trial with psychometric and psychophysiological outcomes. *Angewandte Chemie International Edition*, *6*(11), 951−952.

Jang, S., Kim, J., Kim, S., Hong, J., Kim, S., & Kim, E. (2021). Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit : A development and feasibility/ usability study. *International Journal of Medical Informatics*, *150*, e104440. https://doi.org/10.1016/j.ijmedinf.2021.104440

Johansson, M., Berman, A. H., Sinadinovic, K., Lindner, P., Hermansson, U., & Andréasson, S. (2021). Effects of internet-based cognitive behavioral therapy for harmful alcohol use and alcohol dependence as self-help or with therapist guidance: Three-armed randomized trial. *Journal of Medical Internet Research*, *23*(11), e29666. https://doi.org/10.2196/29666

Kamita, T., Ito, T., Matsumoto, A., Munakata, T., & Inoue, T. (2019). A chatbot system for mental healthcare based on SAT counseling method. *Mobile Information Systems*, *2019*(2), 1−11. https://doi.org/10.1155/2019/9517321

Kang, S. H., & Gratch, J. (2014). Exploring users' social responses to computer counseling interviewers' behavior. *Computers in Human Behavior*, *34*, 120−130.

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., … Cuijpers, P. (2017). Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: A meta-analysis of individual participant data. *JAMA Psychiatry*, *74*(4), 351−359.

Kim, H. C., Cha, M. C., & Ji, Y. G. (2021). The impact of an agent's voice in psychological counseling: Session evaluation and counselor rating. *Applied Sciences*, *11*(7), e2893. https://doi.org/10.3390/app11072893

Labroo, A. A., Dhar, R., & Schwarz, N. (2008). Of frog wines and frowning watches: Semantic priming, perceptual fluency, and brand evaluation. *Journal of Consumer Research*, *34*(6), 819−831.

Latynov, V., & Shepeleva, E. (2020). Applied aspects of the use of algorithms of digital psychometrics. *Psikhologicheskii Zhurnal*, *41*(4), 66−77.

Law, T., Chita-Tegmark, M., Rabb, N., & Scheutz, M. (2022). Examining attachment to robots: Benefits, challenges, and alternatives. *Association for Computing Machinery Transactions on Human-Robot Interaction*, *11*(4), 1−18. https://doi.org/10.1145/3526105

Lederman, R., & D'Alfonso, S. (2019). Insights on the therapeutic alliance in digital mental health. In R. Lederman, & S. D'Alfonso (Eds.), *digitalta2019: Symposium on the Therapeutic Alliance in Digital Mental Health* (pp. 1−3). Melbourne.

Lee, S. K., Kavya, P., & Lasser, S. C. (2021). Social interactions and relationships with an intelligent virtual agent. *International Journal of Human Computer Studies*, *150*, 102608. https://doi.org/10.1016/j.ijhcs.2021.102608

Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human-Computer Interaction*, *36*(10), 930−940.

Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020). "I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. In *Proceedings of Conference on Human Factors in Computing Systems* (pp. 1−12). New York, NY: Association for Computing Machinery.

Leeuwerik, T., Cavanagh, K., & Strauss, C. (2019). Patient adherence to cognitive behavioural therapy for obsessive-compulsive disorder: A systematic review and meta-analysis. *Journal of Anxiety Disorders*, *68*, e102135. https://doi.org/10.1016/j.janxdis.2019.102135

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management*, *40*(3), 191−204.

Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students : A randomized trial of effectiveness. *Internet Interventions*, Advance online publication. https://doi.org/10.1016/j.invent.2022.100495

Luo, X., Bugatti, M., Molina, L., Tilley, J. L., Mahaffey, B., & Gonzalez, A. (2022). Conceptual invariance, trajectories, and outcome associations of working alliance in unguided and guided internet-based psychological interventions: Secondary analysis of a randomized controlled trial. *JMIR Mental Health*, *9*(6), 1−15.

Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, *10*, 39−46.

Mallinckrodt, B., & Jeong, J. (2015). Meta-analysis of client attachment to therapist: Associations with working alliance and client pretherapy attachment. *Psychotherapy*, *52*(1), 134−139.

McGonagle, G., Bucci, S., Varese, F., Raphael, J., & Berry, K. (2021). Is adult attachment associated with engagement with services? A systematic literature review. *Journal of Mental Health*, *30*(5), 607−618.

McGreevey, J. D., Hanson, C. W., & Koppel, R. (2020). Clinical, legal, and ethical aspects of artificial intelligence-assisted conversational agents in health care. *JAMA - Journal of the American Medical Association*, *324*(6), 552−553.

Mikulincer, M., & Shaver, P. R. (2020). Enhancing the “broaden and build” cycle of attachment security in adulthood: From the laboratory to relational contexts and societal systems. *International Journal of Environmental Research and Public Health*, *17*(6), 2054.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, *58*(3), 20−38.

Mori, M., Macdorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, *19*(2), 98−100.

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2021). *That's so embarrassing! When not to design for social presence in human−chatbot interactions*. Paper presented at the meeting of the 42nd International Conference on Information Systems, Austin, USA.

Mrazek, A. J., Mrazek, M. D., Cherolini, C. M., Cloughesy, J. N., Cynman, D. J., Gougis, L. J., … Schooler, J. W. (2019). The future of mindfulness training is digital, and the future is now. *Current Opinion in Psychology*, *28*, 81−86.

Muresan, A., & Pohl, H. (2019). Chats with bots: Balancing imitation and engagement. *In Proceedings of Conference on Human Factors in Computing Systems* (pp. 1−6). New York, NY: Association for Computing Machinery.

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI) -led chatbot services in healthcare : A mixed-methods study. *Digit Health, 5*, 1−12.

Nanjappa, S., Chambers, S., Marcenes, W., Richards, D., & Freeman, R. (2014). A theory led narrative review of one-to-one health interventions: The influence of attachment style and client provider relationship on client adherence. *Health Education Research*, *29*(5), 740−754.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72−78). Boston: Association for Computing Machinery.

Nißen, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., Wangenheim, F. V., & Kowatsch, T. (2022). The effects of health care chatbot personas with different social roles on the client-chatbot bond and usage intentions: Development of a design codebook and web-based study. *Journal of Medical Internet Research*, *24*(4), e32630. https://doi.org/10.2196/32630

Park, D. Y., & Kim, H. (2023). Determinants of intentions to use digital mental healthcare content among university students, faculty, and staff: Motivation, perceived usefulness, perceived ease of use, and parasocial interaction with AI chatbot. *Sustainability*, *15*(1), 872. https://doi.org/10.3390/su15010872

Pedeliento, G., Andreini, D., Bergamaschi, M., & Salo, J. (2016). Brand and product attachment in an industrial context: The effects on brand loyalty. *Industrial Marketing Management*, *53*, 194−206.

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140, 107600. https://doi.org/10.1016/j.chb.2022.107600

Perski, O., Crane, D., Beard, E., & Brown, J. (2019). Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *Digit Health*, *5*, 1−13.

Proudfoot, J., Parker, G., Pavlovic, D. H., Manicavasagar, V., Adler, E., & Whitton, A. (2010). Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. *Journal of Medical Internet Research*, *12*(5), 1−13.

Provoost, S. (2021a). *Embodied conversational agents in internet-based cognitive behavioral therapy for depression bridging the gap between unguided and guided interventions*. Vrije Universiteit Amsterdam.

Provoost, S., Kleiboer, A., Ornelas, J., Bosse, T., Ruwaard, J., Rocha, A., … Riper, H. (2020b). Improving adherence to an online intervention for low mood with a virtual coach: Study protocol of a pilot randomized controlled trial. *Trials*, *21*(1), 1−12.

Puls, H. C., Schmidt, R., Herpertz, S., Zipfel, S., Tuschen-Caffier, B., Friederich, H. C., … Hilbert, A. (2020). Adherence as a predictor of dropout in Internet- based guided self-help for adults with binge-eating disorder and overweight or obesity. *International Journal of Eating Disorders*, *53*(4), 555−563.

Rabb, N., Law, T., Chita-Tegmark, M., & Scheutz, M. (2022). An attachment framework for human-robot interaction. *International Journal of Social Robotics*, *14*(2), 539−559.

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human - Computer Studies*, *151*, 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, *32*(4), 329−342.

Rodrigues, T. R. C., Reijnders, T., de Buisonjé, D. R., Santhanam, P., Kowatsch, T., Janssen, V. R., … Evers, A. W. M. (2021). Human cues in self-help lifestyle interventions : An experimental field study table of contents. *Journal of Medical Internet Research*, Advance online publication. https://doi.org/10.2196/preprints.30057

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*(2), 95−103.

Rubin, A. M. (1983). Television uses and gratifications: The interactions of viewing patterns and motivations. *Journal of Broadcasting & Electronic Media*, *27*(1), 37−51.

Rusbult, C. E., Drigotas, S. M., & Verette, J. (1994). The investment model: An interdependence analysis of commitment processes and relationship maintenance phenomena. In D. J. Canary, & L. Stafford (Eds.), *Communication and relational maintenance* (pp. 115−139). Academic Press.

Sánchez-Franco, M. J., Arenas-Márquez, F. J., & Alonso-Dos-Santos, M. (2021). Using structural topic modelling to predict users' sentiment towards intelligent personal agents. An application for Amazon's echo and Google Home. *Journal of Retailing and Consumer Services*, *63*, 102658. https://doi.org/10.1016/j.jretconser.2021.102658

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875−900.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021a). My chatbot companion: A study of human-chatbot relationships. *International Journal of Human Computer Studies*, *149*, e102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022b). A longitudinal study of human−chatbot relationships. *International Journal of Human Computer Studies*, *168*, 102903. https://doi.org/10.1016/j.ijhcs.2022.102903

Song, S. W., & Shin, M. (2022). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity. *International Journal of Human−Computer Interaction*, Advance online publication. https://doi.org/10.1080/10447318.2022.2121038

Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*, *124*, 106934. https://doi.org/10.1016/j.chb.2021.106934

Spijkerman, M. P. J., Pots, W. T. M., & Bohlmeijer, E. T. (2016). Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. *Clinical Psychology Review*, *45*, 102−114.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645−726.

Sun, S., Lin, D., Goldberg, S., Shen, Z., Chen, P., Qiao, S., … Operario, D. (2021). A mindfulness-based mobile health (mHealth) intervention among psychologically distressed university students in quarantine during the COVID-19 pandemic: A randomized controlled trial. *Journal of Counseling Psychology*, *69*(2), 157−171.

Sundar, S. S., Oeldorf-Hirsch, A., & Garga, A. (2008). A cognitive-heuristics approach to understanding presence in virtual environments. In A. Spagnolli, & L. Gamberini (Eds.), In *Proceedings of the 11th Annual International Workshop on Presence* (pp. 219−228). Padova, Italy: CLEUP Cooperativa Libraria Universitaria Padova.

Tanis, M., & Postmes, T. (2003). Social cues and impression formation in CMC. *Journal of Communication*, *53*(4), 676−693.

Taylor, H., Strauss, C., & Cavanagh, K. (2021). Can a little bit of mindfulness do you good? A systematic review and meta-analyses of unguided mindfulness-based self-help interventions. *Clinical Psychology Review*, *89*, e102078. https://doi.org/10.1016/j.cpr.2021.102078

ter Stal, S., Kramer, L. L., Tabak, M., op den Akker, H., & Hermens, H. (2020). Design features of embodied conversational agents in eHealth: A literature review. *International Journal of Human Computer Studies*, *138*, 102409. https://doi.org/10.1016/j.ijhcs.2020.102409

Tetley, A., Jinks, M., Huband, N., & Howells, K. (2011). A systematic review of measures of therapeutic engagement in psychosocial and psychological treatment. *Journal of Clinical Psychology*, *67*(9), 927−941.

Tong, F., Lederman, R., Alfonso, S. D., Berry, K., & Bucci, S. (2022). Digital therapeutic alliance with fully automated mental health smartphone apps : A narrative review. *Frontiers in Psychiatry*, *13*, 819623. https://doi.org/10.3389/fpsyt.2022.819623

Tversky, A., & Kahnemann, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124−1131.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational

agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry*, *64*(7), 456−464.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, *23*(1), 3−43.

Wanser, S. H., Vitale, K. R., Thielke, L. E., Brubaker, L., & Udell, M. A. R. (2019). Spotlight on the psychological basis of childhood pet attachment and its implications. *Psychology Research and Behavior Management*, *12*, 469−479.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383−388.

Weisel, K. K., Fuhrmann, L. M., Berking, M., Baumeister, H., Cuijpers, P., & Ebert, D. D. (2019). Standalone smartphone apps for mental health: A systematic review and meta-analysis. *Nature Partner Journals Digital Medicine*, *2*(1), 1−10.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907−931.

Xie, C., Wang, Y., & Cheng, Y. (2022). Does artificial intelligence satisfy you? A meta-analysis of user gratification and user satisfaction with AI-Powered chatbots. *International Journal of Human−Computer Interaction*. Advance online publication. https://doi.org/10.1080/10447318.2022.2121458

Xie, T., & Pentina, I. (2022). Attachment theory as a framework to understand relationships with social chatbots: A case study of replika. In *Proceedings of the 55th Hawaii International Conference on System Sciences* (pp. 2046−2055). University of Hawaii Press.Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, *46*(1), 53−93.

Zhu, Y., Wang, R., & Pu, C. (2022). “I am chatbot, your virtual mental health adviser.” What drives citizens’ satisfaction and continuance intention toward mental health chatbots during the COVID-19 pandemic? An empirical study in China. *Digital Health*, *8*, 1−15.

From anthropomorphic attribution to alliance establishment: The effect of human-chatbot relationships on engagement

MO Ran[1], FANG Jiandong[1,2,3], CHANG Baorui[1,2,3]

([1] *Department of Psychology, Faculty of Education, Guangxi Normal University, Guilin* 541006, *China*)

([2] *Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin,* 541004, *China*)

([3] *Guangxi Ethnic Education Development Research Center, Key research base of Humanities and Social Sciences in Guangxi Universities, Guilin* 541006, *China*)

Abstract: AI chatbots can replicate human guidance to improve user engagement and efficacy in internet-based self-help interventions (ISIs), thanks to the rapid development of artificial intelligence (AI) technology. However, the study of chatbots’ mechanisms is still in its early stages. To deepen the rational understanding of this issue, we propose a theoretical model based on the human-computer relationship that adapts to the ISIs situation: chatbots can develop human-chatbot relationships (HCRs) through the four stages of anthropomorphic attribution, utilitarian value judgment, attachment relationship development and the establishment of digital therapeutic alliance (DTA) to improve user engagement. In future research, there is a need to further enrich and evaluate the key HCRs theories, construct chatbots based on the HCRs theory, examine additional variables that affect HCRs, unify operational definitions of engagement, and develop appropriate engagement measurement methods.

Keywords: chatbot, engagement, human-chatbot relationships